

National Library
of Medicine

PubMed

PubMed

Nucleotide

Protein

Genome

Structure

PopSet

Taxonomy

OMIM

Search

PubMed

▼ for

Limits

Preview/Index

History

Clipboard

Go

Clear

Display

Citation

▼

Save

Text

Order

Add to Clipboard

1: *Comput Appl Biosci* 1993 Jun;9(3):299-314

Related Articles 30/3/93

A system for pattern matching applications on biosequences.

Mehldau G, Myers G

Department of Computer Science, University of Arizona, Tucson 85721.

ANREP is a system for finding matches to patterns composed of (i) spacing constraints called 'spacers', and (ii) approximate matches to 'motifs' that are, recursively, patterns composed of 'atomic' symbols. A user specifies such patterns via a declarative, free-format and strongly typed language called A that is presented here in a tutorial style through a series of progressively more complex examples. The sample patterns are for protein and DNA sequences, the application domain for which ANREP was specifically created. ANREP provides a unified framework for almost all previously proposed biosequence patterns and extends them by providing approximate matching, a feature heretofore unavailable except for the limited case of individual sequences. The performance of ANREP is discussed and an appendix gives a concise specification of syntax and semantics. A portable C software package implementing ANREP is available via anonymous remote file transfer.

MeSH Terms:

- Base Sequence*
- Human
- Molecular Sequence Data
- Pattern Recognition*
- Software*
- Support, Non-U.S. Gov't
- Support, U.S. Gov't, P.H.S.
- User-Computer Interface

Grant support:

- R01-4960 PHS

PMID: 8324630